

Please do not distribute. Thanks.

Understanding Urban Dynamics via State-sharing Hidden Markov Model

ABSTRACT

With the ever-increasing urbanization process, systematically modeling people's activities in the urban space is being recognized as a crucial socio-economic task. However, it is extremely challenging due to the scarcity of reliable data and deficiency of suitable methods. The emergence of population-scale urban mobility data sheds new light on it, but recent works on discovering activity patterns from urban mobility data are still limited, especially in terms of concisely and specifically modeling the temporal dynamics of people's urban activities.

To bridge the gap, we present State-sharing Hidden Markov Model (SSHMM), a novel time series modeling method that uncovers urban dynamics with massive urban mobility data. SSHMM models the urban dynamics from two aspects. First, it extracts the urban states from the whole city, which captures the volume of population flows as well as the frequency of each type of Point of Interests (PoIs) visited. Second, it characterizes the urban dynamics of each urban region as the state transition on the shared-states, which reveals distinct daily rhythms of urban activities. We evaluate our method via a large-scale real-life mobility dataset. The results demonstrate that SSHMM learns semantics-rich urban dynamics, which are highly correlated with the functions of the region. Besides, it recovers the urban dynamics in different time slots with an error of 0.0793 when only learning limited states for the whole city, which outperforms the general HMM by 54.2%.

CCS CONCEPTS

- Information systems → Spatial-temporal systems; Data mining;
- Human-centered computing → Empirical studies in ubiquitous and mobile computing;
- Computing methodologies → Machine learning;

KEYWORDS

Urban Computing, Time Series Analysis, Urban Dynamics, Mobility, Hidden Markov Model

ACM Reference Format:

. 2018. Understanding Urban Dynamics via State-sharing Hidden Markov Model. In *Proceedings of ACM Conference (Conference'17)*, Jennifer B. Sartor, Theo D'Hondt, and Wolfgang De Meuter (Eds.). ACM, New York, NY, USA, 11 pages. https://doi.org/10.475/123_4

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

Conference'17, July 2017, Washington, DC, USA

© 2018 Copyright held by the owner/author(s).

ACM ISBN 123-4567-24-567/08/06.

https://doi.org/10.475/123_4

1 INTRODUCTION

The rapid urbanization process has been nurturing large and complex urban systems worldwide. It is estimated that over the next thirty years, more than two-thirds of the people will dwell in modern cities [1]. Therefore, understanding urban dynamics, namely, the temporal patterns of urban activities, is fundamental for tackling the increasingly prominent urban challenges, e.g., excessive energy consumption, air pollution, and traffic congestion [33]. However, citizen's activities in the urban space are extremely complex and highly volatile, which poses challenges to precisely and systematically model temporal urban dynamics. Traditional approaches rely on expensive manual surveys, yet the understanding is still coarse-grained and limited in geographical scope [2].

Fortunately, the advent of the ubiquitous mobile Internet and Location-Based Social Networks (LBSNs) makes it possible to collect population-scale urban mobility data, which sheds new light on this open problem. These datasets contain semantic-rich human mobility information, which includes the timestamps, location coordinates as well as the visited Points of Interest (PoIs). Previous works have demonstrated that the daily movements of citizens can be utilized to infer the functions of urban regions [25, 28], and the patterns of urban activities (e.g., working, resting, commuting, etc.) are closely correlated with urban mobility patterns [24, 29], which indicates the feasibility of leveraging urban mobility data to model urban dynamics.

In this paper, we aim to harness the power of massive urban mobility data to deepen the understanding of urban dynamics. The research problem is non-trivial mainly for three reasons: (1) Urban mobility behaviour is a *noisy* representation of urban activities. Similar urban activities may correspond to slightly different mobility patterns, e.g., central business districts may experience different population flow during the working hours of different days. Therefore, it is difficult to robustly and accurately infer the underlying urban activities from the empirical observation of urban mobility behaviour. (2) The semantic-rich mobility data, i.e., check-in data on LBSNs, is sparse in urban space, especially in sparsely populated areas. The sparse and unevenly distributed mobility data poses significant challenges to extract reliable patterns of urban dynamics for different regions in urban systems. (3) Modern cities are complicated socio-economic systems, where each urban region may possess different urban dynamics due to its unique activities. Therefore, it is challenging to interpret the identified urban dynamics and reveal the underlying mechanisms.

Motivated by these challenges, we propose a novel State-sharing Hidden Markov Model (SSHMM) to reveal the urban dynamics. The key insight of SSHMM is that the mobility behavior of an urban region can be viewed as a probabilistic observation of the underlying urban activity, and similar activities have the similar probability distribution of the observations across different urban regions since they are likely conducted by similar population. Specifically, the model learns the urban dynamics of a region as transitions between

hidden Markov states, where each state maps to a certain urban activity. The corresponding mobility behavior is generated from the hidden state through an emission probability function, which allows the same urban activities to be mapped to slightly different mobility behavior and effectively addresses the problem of noisy mobility data. More importantly, SSHMM facilitates different urban regions to share the same set of hidden states because the similar urban activities in different regions correspond to similar mobility behavior patterns. Therefore, it addresses the challenge of data sparsity by allowing different regions to share parameters, which fully exploits the correlation between different regions. Finally, as a generative model, apart from predicting the mobility behavior of urban regions, SSHMM can also characterize the urban dynamics as hidden state sequences. Based on the identified state sequences, we further design an unsupervised clustering analysis technique, which reveals their correlations with urban functions (i.e., land use) and provides a meaningful interpretation of the urban dynamics.

The contributions of our research are three-fold:

- (1) We propose a novel urban dynamics revealing model SSHMM. It can robustly and accurately infer the underlying urban activities from noisy and sparse urban mobility data by sharing model parameters across different regions. In addition, it achieves qualitative representations of urban dynamics as the transition patterns between urban activities. Compared with previous works, it can model urban dynamic in a concise and probabilistic way.
- (2) We propose an effective and efficient algorithm to infer the parameters of our model. By splitting the long observations into shorter ones and updating the parameters in parallel, we reduce the training time of learning R groups of parameters to that of only one group, where R is the number of regions.
- (3) We evaluate our method using population-scale mobility dataset. The results demonstrate that SSHMM learns meaningful and explainable urban dynamics. Besides, the activity regularities can be recovered with an error of 0.0793 by 100 states, outperforming baseline method by 54.2%. We also achieve RMSE for population flow prediction of 0.195 and *Top3-accuracy* for PoI popularity prediction of 41.4%, which outperforms the baseline method by 16% and 8%, respectively.

The rest of this paper is organized as follows. After reviewing the related works in Section 2, we formulate the problem and provide an overview of our system in Section 3. Motivated by the challenges, we introduce our modeling method in details in Section 4. After that, we apply our system to a real-world mobility dataset and give the evaluations in Section 5. We discuss the implications and future works in Section 6 and conclude our paper in Section 7.

2 RELATED WORK

Urban dynamics modeling. The development of the city has witnessed a great series of studies on urban dynamic problems. Urban dynamics, generally defined as how sociological indicators (e.g., the population, the land use) change over time [7], can be divided into two aspects. One is to investigate the urbanization and sustained economic growth [16] via the dataset with a long period, while another focuses on human daily activity rhythms in the city with

more fine-grained time granularity [2, 18], which is more relevant to our research. Zhang et al. [33] used the geo-tagged social data to model urban activities with more attention paid to temporal dimension, which demonstrated that the activity volume of an area is not uniformly distributed across time and different areas have different activity volume temporal distributions. Sofiane et al. [2] built activity time series for different cities, and different neighborhoods within the same city to identify the different dynamic patterns taking place via the geo-tagged data from Twitter. By clustering the activity time series, they found that close neighborhoods tend to share similar rhythms. Miranda et al. [18] captured the spatio-temporal activity in a city across multiple temporal resolutions and visualized different activity levels in different time.

In this problem, we regard urban activities as time series and aim to reveal the daily regularity hidden in them. Different from the existing works based on statistical analysis [2] and data visualization [18], we propose a specific model achieving understanding and prediction at the same time.

Urban computing with mobility data. Mobility is the fundamental activity in people's daily life [9], thus understanding human mobility patterns with its underlying mechanisms is vital for urban computing to solve massive urban issues [34]. The availability of large-scale mobility data, especially the geo-tagged trajectories such as check-ins, sheds new light on data-driven urban computing methods. Zhang et al. [29, 30] utilized the geo-tagged social media to model people's activities in the urban space via representative learning, which gives deep insight into the relationship between urban dynamics and human mobility. Moreover, by utilizing the aggregated mobility in the city areas, Yuan et al. [28] proposed an LDA model to detect the existence of different functional regions in a city through the GPS trajectory and PoI datasets. Yao et al. [25] presented a city zone embedding framework using human mobility patterns to present urban functions as distributed and low-dimensional vectors via massive taxi trajectories. Different from the above methods, our model uses hidden state learning from human mobility and semantic check-ins to represent urban dynamics, and we can infer the urban functions by different dynamic patterns.

Hidden Markov Model and its application. Hidden Markov Model (HMM) is a statistical Markov model in which the system being modeled is assumed to be a Markov process with unobserved (i.e., hidden) states [21], which has been widely used in time series analysis and prediction in the past years. In HMM, model mixture and parameter sharing are very helpful to deal with increasingly complex tasks [10]. One of the well-known mixture methods is the Gaussian mixture model based HMM (GMM-HMM). GMM is parametric probability density function represented as a weighted sum of Gaussian component densities, while GMM-HMM assumes that a set of Gaussian components can represent a distribution based on the spectral envelope [8]. The second example is shared-distribution HMM, where clustering is carried out at the distribution level for parameters sharing and output distributions are shared with each other if they exhibit acoustic similarity [12]. Another model tied-mixture HMM uses both mixture and parameter sharing, which belongs to semi-continuous HMM. It is recognized as a useful complexity reduction method, because of its ability to maintain modeling accuracy of large-mixture probability density functions (PDF) by enforcing PDF sharing [3, 11, 15].

SSHMM is also a kind of parameters sharing HMM. While different from previous works, we design it to automatically learn a set of states for the continuous observations and no following clustering is conducted to force the parameters shared. Both based on the Gaussian emission function, we use multi-dimentional Gaussian instead of GMM. In addition, though HMM has been successfully applied into the topic of mobility modeling, most of the work concentrated on individual mobility prediction [17, 31, 36], and we are the first to apply it in urban dynamics modeling.

3 OVERVIEW

3.1 Problem Description

HMM is a generative model, which assumes that the observation sequence is generated by a hidden state sequence. To apply this model to urban dynamics problem, we regard human aggregated activities in different regions as time series. More specifically, we extract mobility behaviours in different time slots as observation sequences, and we aim to reveal urban dynamics in terms of human daily life rhythms by the corresponding hidden state sequence. Before formally define our investigated problem, we give the definition of mobility behaviour observation as follows,

Definition 1 (Mobility Behaviour Observation). *The mobility behaviour sequences of region r is a time-ordered sequence $O_r = [O_{r,1}, O_{r,2}, \dots, O_{r,N}]$, where $O_{r,n}$ is a tuple of length L , standing for the observation in n -th time slot. It contains two parts: (1) $\{o_{r,n,1}, o_{r,n,2}, o_{r,n,3}\}$ denotes the number of people who are arriving at, leaving from and staying in this region in this time slot. (2) $\{o_{r,n,4}, o_{r,n,5}, \dots, o_{r,n,L}\}$ denotes the check-in frequency of different categories of Point of Interests (PoIs).*

The motivation of choosing these two types of features as the observation is very intuitive. On the one hand, how many people leave from, arrive at and stay in each region, represents the daily commute pattern and activity level, is crucial for urban dynamics modeling [4, 27]. On the other hand, since dynamic pattern has close relation with the land use, while the PoIs are static, therefore, we utilize the check-ins, which reflects people's dynamic demand for a region's function [35]. As a consequence, these two aspects can give deep insights into urban activities.

Now, we formulate the urban dynamics understanding problem. Given the mobility behaviour observations of a group of regions in the city, we aim to (1) discover hidden states represented by the volume of population and the visited frequency of different PoIs to understand the intrinsic life modes in the city; (2) reveal urban dynamics represented by hidden state sequences to understand the daily life rhythms and their correlation with urban functions; (3) characterize dynamic regularity by learning state transition probability to achieve dynamic prediction.

3.2 Model Description

Definition 2.1 (SSHMM). *HMM is parameterized by two parts, one is the state transition parameter characterizing how the states transit and the other is the state emission parameter characterizing how the observation generated by the state. SSHMM contains R groups of HMMs for R groups of observation sequences. However, these HMMs share the same emission parameters, which means all the observations are generated by the same set of hidden states.*

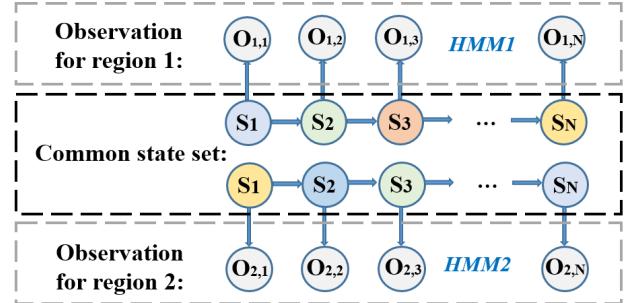


Figure 1: An illustration of the SSHMM, where two kinds of dynamics for region 1 and 2 are generated from the same set of hidden states and each state is presented by a unique color.

Definition 2.2 (Hidden state set). *We define the hidden state set including K hidden states as $S = \{s_1, s_2, \dots, s_K\}$, where each state describes L dimensional features.*

Our model is based on the general assumptions of HMM, where each observation $O_{r,n}$ is generated from one hidden state s_n , and the n -th hidden state s_n merely depends on the previous state s_{n-1} . In our problem, we use one HMM denoted by θ_r to model the dynamics of r -th region, therefore we need to learn R groups of transition parameters for R regions in the city. Figure 1 gives the illustration of our model. There are two observation sequences for region 1 and 2, so we learn two HMMs for them respectively. These observations are generated by the same set of states, where each state is presented by a unique color. These states appear in different time slots, which reveals different dynamics of these two regions.

The key insights of sharing states are two aspect. On the one hand, we investigate the dynamics of the city, where life modes (i.e., sleeping, working, etc.) is limited and similarities in different regions exist objectively. Therefore, it is reasonable to use common state set and it also makes the following dynamics analysis through state sequences more easy. One the other hand, massive mobility data especially the semantic check-in data is sparse and mobility behaviour is a noisy representation of urban activities, which indicates that by sharing states we can achieve more robust and accurate models.

3.3 System Overview

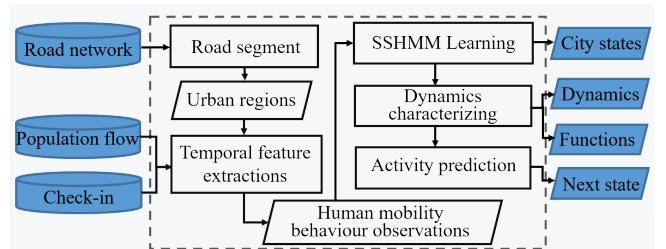


Figure 2: The framework of our system.

Figure 2 summarizes our system framework. We first apply a map segment method [28] to obtain the geographical boundaries of each region formed by the road network. Then we extract the

temporal feature, i.e., the population flow and check-in frequency in each time slot for each region and normalize them as mobility behaviour observations. After that, we learn SSHMM to discover the city states and reveal the dynamics as well as functions. Based on these characterized dynamics, we make city state predictions.

4 METHODOLOGY

In this section, we first introduce the data pre-process and describe our proposed SSHMM method in details. Then, we discuss how to find the regions with similar dynamics with similar functions.

4.1 Data Pre-processing

We first adopt the map segment method [28] to obtain the geographical boundaries of each region formed by the road network. In view of the urban mobility patterns and life-styles, those regions are used as an unit to reveal the dynamics instead of simply dividing the city into grids.

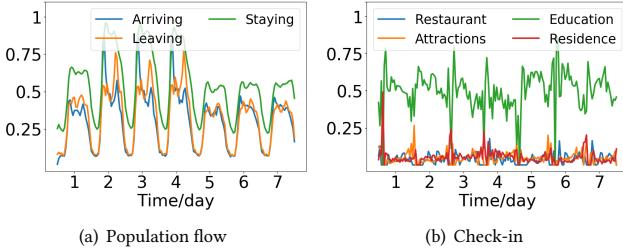


Figure 3: Normalized mobility behaviour observations for Tsinghua University, Beijing from April 1st to 7th, 2018.

To share the states in the city, we normalize mobility behaviour observations to eliminate the problem regarding the difference of population between regions. For mobility, i.e., the number of arriving, leaving and staying, we directly conduct min-max normalization for each region over different time slots respectively. For check-ins, i.e., the visit frequency of POIs, we first compute the TF-IDF weights based on the region-POI matrix in each time slot. TF-IDF [20], short for term frequency-inverse document frequency, is a numerical statistic that is intended to reflect how important a word is to a document in a collection or corpus. Here, we use the TF-IDF weight to indicate how popular a POI category is in a certain time slot. Then, for each region we conduct the min-max normalization on the TF-IDF weights over different time slots. After normalization, each dimension of the mobility behaviour observations is rescaled to 0–1. Figure 3 gives an example of the normalized mobility behaviour observations. From Figure 3 we can observe that both the volume of population flow and the check-in frequency change from morning to night every day.

After pre-processing, we regard the aggregated activities in the city as normalized time series, which is denoted by $O = \{O_1, O_2, \dots, O_r, \dots, O_R\}$ as **Definition 1**.

4.2 Model Definition

In this section, we give the basic assumptions and the formula definition. When generating $o_{r,n}$ from s_n , we assume that the emission probability is Gaussian distribution, i.e., $p(o_{r,n,l}|s_n) = N(o_{r,n,l}|\mu_n, \sigma_n)$. The reason to chose the Gaussian is intuitive: μ_n

describes the fundamental feature of s_n , while the corresponding observations generated by s_n is slightly different from the mean μ_n with the difference controlled by the variance σ_n . To make the following parts more readable, we consistently use the symbols for description as shown in Table 1.

Symbol	Description
r, R	r -th region, number of regions
n, N	n -th time slot, number of time slots
l, L	l -th feature, dimension of features
o, O	Observation in one time slot, observation sequence
s, S	Hidden state in one time slot, hidden state set
θ	HMM parameter set, including π, A, μ, σ
π	Hidden state initial probability
A	Hidden state transition probability
μ, σ	Mean and variance of the Gaussian distribution

Table 1: Notation and Description.

In summary, we build the SSHMM parameterized by R groups of parameters $\theta = \{\theta_1, \theta_2, \dots, \theta_r, \dots, \theta_R\}$ with $\theta_r = \{\pi_r, A_r, \mu, \sigma\}$ denoted for r -th region, where

- (1) $\pi_r \in \mathbb{R}^{K \times 1}$ denotes the initial distribution over K hidden states, i.e., $\pi_{r,k} = p(s_1 = k) (1 \leq k \leq K)$;
- (2) $A_r \in \mathbb{R}^{K \times K}$ denotes the transition probabilities among the K hidden states. If $(n-1)$ -th state is $s_{n-1} = j$, then the probability for n -th state s_n to be k is given by $A_{r,j,k}$, i.e., $p(s_n = k | s_{n-1} = j) = A_{r,j,k}$;
- (3) $\mu, \sigma \in \mathbb{R}^{K \times L}$ denotes the mean and variance of observation probability, i.e., $p(O_{r,n}|s_n = k) = \prod_{l=1}^L \frac{1}{\sqrt{2\pi\sigma_{k,l}}} \exp\left(-\frac{(o_{r,n,l} - \mu_{k,l})^2}{2\sigma_{k,l}}\right)$.

It is worth noting that we do not use the subscript r to distinguish μ, σ of different HMMs, because in SSHMM they are determined by the common state set.

4.3 Model learning

To reveal urban dynamics by the hidden states sequence, the first step is to learn the model parameters by maximizing the probability of the observation sequences given the model. Since SSHMM is a novel model proposed by us, how to learn the model parameters reasonably and effectively and how to apply it in the real-world large-scale dataset are two difficulties we face.

4.3.1 Parameter Inference. To infer the parameters, we use Expectation-Maximization method (EM) [6] as the solution. Since SSHMM shares the same set of hidden states, the existing famous Baum-Welch algorithm [23] cannot be applied directly. To address this, we give a new parameter derivation process, and in order to show it more clearly, we first give some definitions and lemmas as follows.

Definition 3.1 (Log likelihood $L(\theta)$). *Let the log likelihood for r -th region $L(\theta_r) = \ln p(O_r|\theta_r) = \ln \sum_S p(O_r|S, \theta_r)p(S|\theta_r)$. Therefore, the log likelihood for all observation sequences $L(\theta)$ can be derived as follows,*

$$L(\theta) = \ln p(O|\theta) = \ln \prod_{r=1}^R p(O_r|\theta_r) = \sum_{r=1}^R \ln p(O_r|\theta_r) \\ = \sum_{r=1}^R L(\theta_r). \quad (1)$$

Definition 3.2 (Q-function $Q(\theta, \theta^t)$). Let the Q-function for r -th region $Q(\theta_r, \theta_r^t) = \sum_S p(S|O_r; \theta_r^t) \ln p(O_r, S|\theta_r)$, where θ_r means the old parameter. Therefore, the total Q-function for regions $Q(\theta, \theta^t)$ can be derived as follows,

$$\begin{aligned} Q(\theta, \theta^t) &= \sum_S \ln p(O|S, \theta) p(S|O, \theta^t) \\ &= \sum_S \ln \prod_{r=1}^R p(O_r|S, \theta_r) p(S|O_r, \theta_r^t) \\ &= \sum_{r=1}^R \sum_S \ln p(O_r|S, \theta_r) p(S|O_r, \theta_r^t) = \sum_{r=1}^R Q(\theta_r, \theta_r^t). \end{aligned} \quad (2)$$

Lemma 1. To achieve the maximized log likelihood, parameters can be updated by maximizing $Q(\theta, \theta^t)$ step by step.

Proof. For each HMM parameterized by θ_r , Welch have proven that maximization of $Q(\theta_r, \theta_r^t)$ leads to increased likelihood [21], i.e.,

$$\max_{\theta_r^t} Q(\theta_r, \theta_r^t) \Rightarrow p(O_r|\theta_r^t) \geq p(O_r|\theta_r). \quad (3)$$

Eventually the likelihood converges to a critical point. From Eq. 1 and 2, we yield that

$$Q(\theta, \theta^t) = \sum_{r=1}^R Q(\theta_r, \theta_r^t), L(\theta) = \sum_{r=1}^R L(\theta_r). \quad (4)$$

Take the summary of the Q-function for all regions, the following result from Eq. 3 could be obtained as,

$$\sum_{r=1}^R \max_{\theta_r^t} Q(\theta_r, \theta_r^t) \Rightarrow \sum_{r=1}^R p(O_r|\theta_r^t) \geq p(O_r|\theta_r). \quad (5)$$

Combine Eq. 4 with Eq. 5, we can further derive the relationship between Q-function and log likelihood as follows,

$$\max_{\theta^t} Q(\theta, \theta^t) \Rightarrow p(O|\theta^t) \geq p(O|\theta). \quad (6)$$

This is to say, by maximizing $Q(\theta, \theta^t)$, the likelihood can converge to its maximum.

Since $p(O_r|S, \theta_r^t)$ can be calculated as follows,

$$\begin{aligned} p(O_r|S, \theta_r^t) &= \pi_{r,k} p(O_{r,1}|s_{r,1}) \cdot p(s_{r,2}|s_{r,1}) p(O_{r,2}|s_{r,2}) \\ &\quad \cdots (s_{r,N}|s_{r,N-1}) p(O_{r,N}|s_{r,N}). \end{aligned} \quad (7)$$

Put Eq. 7 into the Eq. 2, the Q-function can be further unfolded as follows,

$$\begin{aligned} Q(\theta, \theta^t) &= \sum_{r=1}^R \sum_S p(S|O_r, \theta_r^t) \ln \pi_{r,k} \\ &\quad + \sum_{r=1}^R \sum_S \sum_{n=1}^{N-1} p(S|O_r, \theta_r^t) \ln p(s_{n+1}|s_n) \\ &\quad + \sum_{r=1}^R \sum_S \sum_{n=1}^N p(S|O_r, \theta_r^t) \ln p(O_{r,n}|s_{r,n}), \end{aligned} \quad (8)$$

From Eq. 8, we can observe that the optimization parameters π, A (i.e., $p(s_{n+1}|s_n)$) and $\{\mu, \sigma\}$ (i.e., $p(O_{r,n}|s_{r,n})$) appear independently in the upper formula, so it is reasonable to maximize each item separately. As a consequence, the parameters can be optimized by Eq. 9 and more detailed derivations for $\alpha, \beta, \gamma, \xi$ can be found in APPENDIX.

$$\begin{aligned} \pi_{r,k}^{(t+1)} &= \gamma(s_{r,1}^k), \\ A_{r,j,k}^{(t+1)} &= \frac{1}{\Xi_j} \sum_{n=2}^N \xi(s_{r,n-1}^j, s_{r,n}^k), \\ \mu_{k,l}^{(t+1)} &= \frac{1}{\Gamma_K} \sum_{r=1}^R \sum_{n=1}^N \gamma(s_{r,n}^k) o_{r,n,l}, \\ \sigma_{k,l}^{(t+1)} &= \frac{1}{\Gamma_K} \sum_{r=1}^R \sum_{n=1}^N \gamma(s_{r,n}^k) (o_{r,n,l} - \mu_{k,l}^{(t+1)})^2, \end{aligned} \quad (9)$$

$$\text{where } \Gamma_K = \sum_{r=1}^R \sum_{n=1}^N \gamma(s_{r,n}^k), \Xi_j = \sum_{n=2}^N \sum_{i=1}^K \xi(s_{r,n-1}^j, s_{r,n}^i).$$

Eq. 9 is consistent with our intuition of **Definition 2.1**. $\pi_{r,k}$ and $A_{r,j,k}$ as state transition parameters are only depended on the observations of the corresponding region r , while $\mu_{k,l}, \sigma_{k,l}$ as emission parameters are depended on the observations of all regions because the state set is shared.

4.3.2 Learning algorithm. The derivation of parameters has been obtained above. However, the process could be inefficient when applied into large-scale datasets due to two aspects. First is the observation sequence length N . The large dataset with a long period leads the computation of α and β complicated as Eq. 15 shows. What's worse, the result may even exceed the max value of the floating-point numbers that the computer can store after several rounds of multiplication. Second is about total number of states K . The time complexity of parameter inference is $O(RNLK^2)$, where R is the number of regions, N is the length of observation sequence, L is the dimension of each observation and K is the number of states. The time is mainly costed by the calculation of ξ as Eq. 16, the computational complexity of which is quadratic in K , rendering it inefficient for large K . However, if the time period is short or K is not large enough, the model is unable to capture all the dynamic patterns in the city, thus reducing the representational ability of the model.

Algorithm 1: SSHMM parameter learning.

Input: Observations $O = \{O_1, O_2, \dots, O_R\}$, Maximum Iterations $MaxIter$;

Output: $\theta_r = \{\pi_r, A_r, \mu, \sigma\} \forall 1 \leq r \leq R$;

Procedure:

Initialization: $t = 0$, initial $\pi_{r,k}^{(0)} = 1/K, A_{r,j,k}^{(0)} = 1/K$, $\mu_{k,l}^{(0)} = \text{random}(0, 1), \sigma_{k,l}^{(0)} = \text{random}(0, 0.1)$, $\forall 1 \leq j, k \leq K, 1 \leq l \leq L$.

while $t < MaxIter$ **do**

for $w = 1, 2, \dots, W$ **do**

E-step: $\forall 1 \leq r \leq R$, calculate $\alpha(s_{r,n})^{(t+1)}, \beta(s_{r,n})^{(t+1)}$, $\gamma(s_{r,n})^{(t+1)}, \xi(s_{r,n})^{(t+1)}$ **in parallel** based on old parameters $\theta_r^{(t)}$ utilizing the w -th subsequence of O_r .

M-step: $\forall 1 \leq r \leq R$, update $\pi_{r,k}^{(t+1)}$ and $A_{r,j,k}^{(t+1)}$ **in parallel**.

Update $\mu_{k,l}^{(t+1)}, \sigma_{k,l}^{(t+1)}$ utilizing $\gamma(s_{r,n})^{(t+1)}, \xi(s_{r,n})^{(t+1)}$ $\forall 1 \leq r \leq R$.

Update t : $t = t + 1$

To overcome these problems, we first split the long observation sequences into several shorter sub-sequences. In each round, we

use only one sub-sequence. By doing this, we can not only utilize all the data for parameter learning but also avoid float-point number exceeding. Second, from (9), we can observe that the updating of $\pi_{r,k}$ and $A_{r,j,k}$ for region r is independent. Thus, we update them in parallel, which cuts down the training time remarkably, as the time complexity has been reduced to $O(NLK^2)$ approximately.

The detailed procedure is shown in Algorithm 1. In each iteration, we run EM-steps for W rounds, and in each round, we fed the subsequences of length N into the model. After calculating the $\alpha(s_{r,n})^{(t+1)}$, $\beta(s_{r,n})^{(t+1)}$, $\gamma(s_{r,n})^{(t+1)}$, $\xi(s_{r,n})^{(t+1)}$, $\pi_{r,k}^{(t+1)}$ and $A_{r,j,k}^{(t+1)}$ in parallel for all the regions in each round, we update the state parameters $\mu_{k,l}^{(t+1)}$, $\sigma_{k,l}^{(t+1)}$ until the convergence is realized.

4.4 Dynamics Revealing and Functions Inferring

With the model $\theta_r = \{\pi_r, A_r, \mu, \sigma\}$ obtained from above algorithm, the state sequences can be decoded by Viterbi algorithm [22], which can further represent the dynamics of that state. Recall that regions with similar dynamic patterns could have similar functions, we match the dynamics with the functions by clustering the state sequence of different regions via k-medoids algorithm [13]. In this clustering algorithm, we define the distance of two sequences as the sum of the Euclidean distance of the mean of the two corresponding states. The specific calculation is as follows,

Definition 4.1 (The distance between corresponding states $d(s_{i,n}, s_{j,n})$). $d(s_{i,n}, s_{j,n})$ is the distance between the state of i -th and j -th region in n -th time slot. It is the Euclidean distance between the mean value vector of Gaussian distribution of these two states,

$$d(s_{i,n}, s_{j,n}) = \sqrt{\sum_{l=1}^L (\mu_{i,n,l} - \mu_{j,n,l})^2}, \quad (10)$$

where $\mu_{i,n,l}, \mu_{j,n,l}$ is the mean of l -th dimension of $s_{i,n}, s_{j,n}$, respectively.

Definition 4.2 (The distance between state sequences $D(S_i, S_j)$). $D(S_i, S_j)$ is the distance between the state sequence of i -th and j -th region. It is the average distance of all the corresponding states,

$$D(S_i, S_j) = \frac{\sum_{n=1}^N d(s_{i,n}, s_{j,n})}{N}, \quad (11)$$

where $\mu_{i,n,l}, \mu_{j,n,l}$ is the mean of l -th dimension of $s_{i,n}, s_{j,n}$, respectively.

We adopt *Davies-Bouldin index* (DBI) [5] to determine the number of clusters, which reflects the ratio between inter-cluster distance and inter-cluster distance. A smaller DBI usually indicates a more effective division. Finally, similar dynamics in each cluster would present the same kind of function in the city.

4.5 Dynamics Prediction

SSHMM also enables the ability of prediction. Through Viterbi algorithm, we can identify the last state of the region based on the observation. With the advantage of a probabilistic model, we can infer the state that the region will be in the next time slot. More specifically, we first obtain the current state of the region according

to the current observation, then predict the next state according to the latest state by maximizing the transition probability.

Formally, given the current observation sequence of r -th region denoted by $O_{r,1} O_{r,2}, \dots, O_{r,n}$, we decode its corresponding hidden state sequence as $s_{r,1} s_{r,2}, \dots, s_{r,n}$. If the latest state $s_{r,n}$ is i -th state in the state set, then the state in the next time slot can be predicted as

$$s_{r,n+1} = \arg \max_{1 \leq j \leq K} A_{r,i,j}. \quad (12)$$

This is to say, we can achieve the prediction for the volume of population flows and the percentage of PoIs visited in the next time slot by utilizing the mean value μ, σ of the identified next state.

5 EXPERIMENTS

In this section, we empirically evaluate the performance of SSHMM. We first introduce the large-scale dataset we utilize and the system parameters we set, and then we show the results we achieve.

5.1 Data

The mobility dataset was collected by collaborating with *Tencent*¹, one of the largest Internet integrated service providers in China. Among their services, there are two very popular ones: location service provided by *Tencent Map*² and social service provided by *Wechat*³. Whenever the user's mobile phone sends a location request by all of the services provided by *Tencent*, her GPS location and request timestamp could be recorded. In this way, we are able to collect users' trajectories. If the request comes from *Wechat Moment*, where the user shares her status with showing the current location, then besides the GPS location, the PoI that the user checks-in is also available.



Figure 4: Road network that divides the downtown area of Beijing into 665 regions.

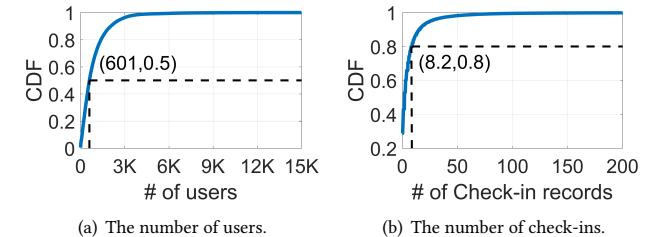


Figure 5: The statics of dataset.

The collected dataset covers near 2 million users in Beijing, China with the duration of one month from April 1st to 30th, 2018. We divide the check-in PoIs into nine categories: *Company*, *Agency*,

¹Tencent Incorporation. <https://www.tencent.com/en-us/index.html>

²Tencent Incorporation. <http://lbs.qq.com/guides/startup.html>

³Tencent Wechat. <https://www.wechat.com/en>

Shopping, Service, Entertainment, Attractions, Education and Residence. We also crawl the road network from Map service and divide the downtown area in Beijing into 665 non-overlapping regions. The regions we select are as shown in Figure 4. We have counted the number of users as well as check-ins in each region and each half an hour, the Cumulative Distribution Functions (CDF) of which are as shown in Figure 5. From these two figures we can observe that our sample has covers a large amount of the population as they are more than 600 users for more than 50% of the cases, while the semantic check-ins is rather sparsity as only 20% of the cases with more than 8 records.

Ethics. It is worth noting that to protect user privacy, all data is anonymous and stored in Tencent offline servers. We pre-process the data under their overseeing and only take the aggregated results of different regions for further analysis. Our research has been reviewed and approved by both the company and our local university institutional board.

5.2 System Settings

5.2.1 Data Usage. In the experiments, we divide the dataset into two parts. We utilize 21 days of data to generate the mobility behaviour observations to train our model and use the rest for prediction evaluation. We set the length of time slot to 1 hour. The observation in each time slot is 12-dimensional, including 3-dimensional population flow volume and 9-dimensional check-in frequency.

5.2.2 Model evaluation. To evaluate the effectiveness of our model and the inferred parameters, we utilize the obtained states to recover the observations by concatenating the mean value of corresponding state in the hidden state sequence in chronological order. We adopt ϵ (see Definition 5), the error between all of the observations and the activities recovered by the mean value of the corresponding hidden state as an evaluation metric.

Definition 5 (Recovering Error ϵ). Given the observation sequence of r -th region $[O_{r,1}, O_{r,2}, \dots, O_{r,N}]$ and the corresponding state sequence $[s_{r,1}, s_{r,2}, \dots, s_{r,N}]$, then the raw observation and the recovered observation in n -th time slot is $\{o_{r,n,1}, o_{r,n,2}, \dots, o_{r,n,L}\}$ and $\{\mu_{n,1}, \mu_{n,2}, \dots, \mu_{n,L}\}$, respectively. Therefore, it is defined as

$$\epsilon = \sqrt{\frac{\sum_{r=1}^R \sum_{n=1}^N \sum_{l=1}^L (\mu_{n,l} - o_{r,n,l})^2}{R \times N \times L}}. \quad (13)$$

To determine the number of hidden states and the length of time slots for training in each round, we try different values for comparison. As shown in Figure 6(a), we can observe that when the number of state K increases, ϵ decreases while the training time increases rapidly. Thus, we set $K = 100$ as a trade-off between model complexity and accuracy. As Figure 6(b) show, we find the length of time window has almost no impact on the training time and ϵ , which is in line with our expectations. We set is as 24, i.e., training the model day by day.

To demonstrate that the dynamics are highly related to urban functions and our model is able to infer functions, we manually label several regions for dynamics validation. Named by their landmarks, the regions with different function types are as shown in Table 2.

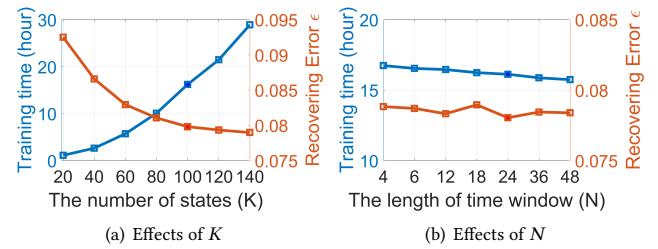


Figure 6: Effects of parameters.

Region Name	Functions
Tsinghua University	Education
Peking University	Education
Tian'anmen	Tourist attraction
Xidan	Shopping & Business
Wangjing	Residence
Zhongguancun Software Park	Company

Table 2: Typical regions with different functions.

5.2.3 Prediction Performance. In terms of prediction, we utilize the rest 9 days' data for evaluation. We compare the mean value of the next state predicted with the ground-truth observations, and the baseline is general HMM. For mobility prediction, we adopt *RMSE* as the metric, which is defined as the average root-mean-square error between the normalized population flow of predicted and the ground truth [32]. For check-ins, we adopt *TopM-accuracy* that reflects the average accuracy on topN frequently-visited PoI prediction of all regions [26]. A lower *RMSE* or a higher *TopM-accuracy* indicates better prediction performance. These two metrics are calculated as follows.

Definition 5.2 (RMSE and TopM-accuracy). Given the ground-truth observation sequence of r -th region $[O_{r,1}, O_{r,2}, \dots, O_{r,N}]$ and the corresponding prediction $[P_{r,1}, P_{r,2}, \dots, P_{r,N}]$, then the ground truth and the prediction in n -th time slot both contains two parts as Definition 1: (1) $\{o_{r,n,1}, o_{r,n,2}, o_{r,n,3}\}$ and $\{p_{r,n,1}, p_{r,n,2}, p_{r,n,3}\}$ denotes the number of arriving, leaving and staying; (2) $\{o_{r,n,4}, \dots, o_{r,n,L}\}$ and $\{p_{r,n,4}, \dots, p_{r,n,L}\}$ denotes the check-in frequency of different categories of PoIs. By sorting the PoIs with their frequency in a descending order and keep the first M PoIs, we obtain the raw TopM list and predicted TopM list denoted by $V_{r,n}^g$ and $V_{r,n}^p$ respectively. Above all, the metrics are defined as follows,

$$RMSE = \sqrt{\frac{\sum_{r=1}^R \sum_{n=1}^N (p_{r,n,l} - o_{r,n,l})^2}{R \times N}}, \quad (14)$$

$$TopM\text{-accuracy} = \frac{\sum_{r=1}^R \sum_{n=1}^N (|V_{r,n}^g \cap V_{r,n}^p| / M)}{R \times N},$$

5.3 Results

5.3.1 Model Effectiveness. To show the representative ability for human activities of our model, we first provide a case study of *Tsinghua University* and then give the overall evaluations. We show the recovering results of *Tsinghua University* of one week as examples in Figure 8, which exhibits a high similarity with the original observations compared with Figure 3. The PoI visited frequencies

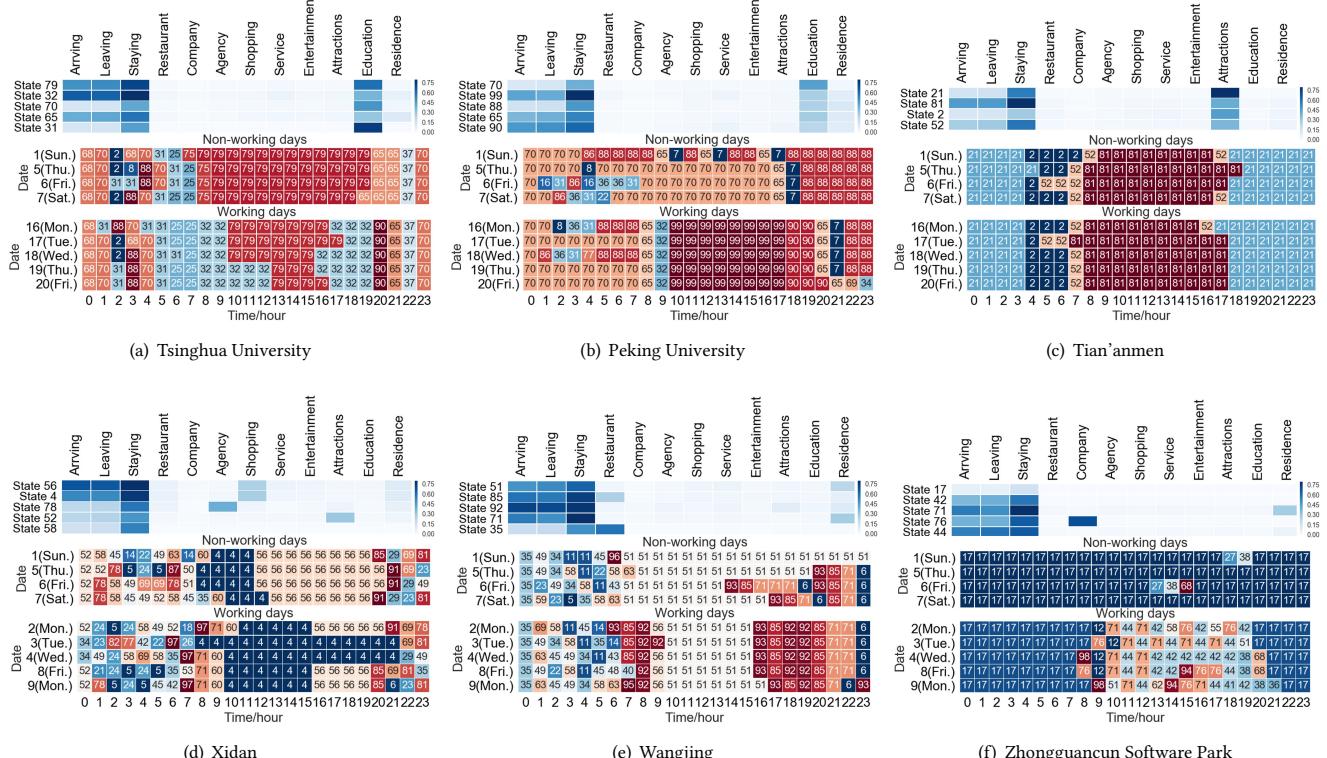


Figure 7: Visualization for representative states and state sequences for different regions, where each row in the state sequences heatmap exhibits the state transition process of one day with the number indicating the corresponding state.

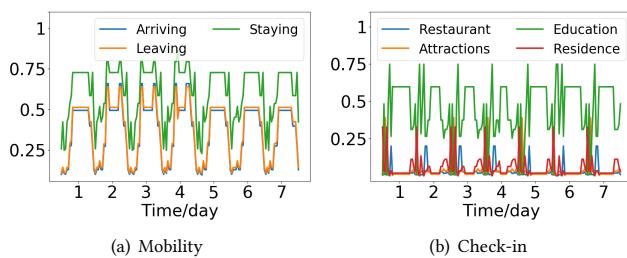


Figure 8: Recovered activities for Tsinghua University.

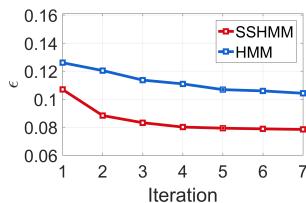


Figure 9: Model Effectiveness comparison.

in original observations are volatile due to the sparsity of check-ins while the recovered results are more regular. Since the mobility is smooth, the recovered results can still capture the peak as well as valley and the difference between working day and non-working day. Figure 9 shows the relationship between recovering error ϵ

and the number of iterations. We compare it with the result of training independent HMM for each region with the same number of states. From Figure 9, we can observe that as the number of iterations increases, the recovering error ϵ decreases first, then tends to remain unchanged. Furthermore, our representative ability for human activities by a limited number of states is better than HMM: When the model converges, our ϵ is 0.0793, which outperforms HMM by 54.2%.

5.3.2 Visualization of the Hidden States and Urban Dynamics. In order to demonstrate the ability of our model in discovering hidden states and revealing urban dynamics, we give a series of special examples and detailed explanations. Figure 7 visualizes the results for the regions in Table 2. We first plot the mean value of the states with frequent occurrence, then show the state transition process in working day and non-working day, respectively. In China, besides the normal weekends, April 5th (Thursday), 6th (Friday) and 7th (Saturday) are the Qingming Festival, which thus belongs to the non-working day, while April 8th (Sunday) is a working day.

First, we look into the discovered states as shown in Figure 7. Each state has the semantics of two aspects: (1) How dense the population flow is. For example, **state 32** presents a large volume of flows and high density of populations, **state 21** presents a small volume of flows while a high density of populations, and **state 17** present a small volume of flows and low density of populations. (2) How frequent the different PoIs are visited. For example, **state**

31 indicates the most frequently visited PoI is *Education*, while **state 21** indicates the most frequently visited PoI is *Attractions*. As shown in Figure 7(a), this region have the **state 79** during the day as *Tsinghua University* occupies most of this region. Likewise, *Peking University* has **state 99** and *Tian'anmen* has **state 81**.

Actually, these two aspects are correlated. By combining their semantics, we can infer the activity level and lifestyle of the region. For example, **state 79** shows active school status while **state 70** shows quiet school status. The rest states are also rich in semantics, but we are not allowed to introduce them all due to the space limitation. However, these high representative states indicating different activity levels and lifestyles, are sufficient to demonstrate the ability of SSHMM to model urban activities through mobility.

Then we discuss the dynamics represented by the state transition processes. Take the dynamics of *Tsinghua University* again as an example. As shown in Figure 7(a), during the night, there are fewer people than the day as **state 70**, and **31** have a smaller mean value of staying than that of **state 79**. Besides, in working days, there is a sudden increase in crowd flow as **state 32** appear in 8:00-9:00 and 17:00-19:00. The transition from **state 70** to **32**, from **32** to **79** in working days reveals the dynamics that only students live in the region at night and many teachers go to school in the morning, causing the population denser than night. Compared with 7(a) and (b), 7(c) shows that both in working day and non-working day, the density of population is consistently high and the PoI visited most frequently is *Attractions*, as *Tian'anmen* is one of the most famous tourist spots in China. Another interesting finding is that for shopping and residential areas, as Figure 7(d) and (e) shown, non-working days are more prosperous and lively than working days, but working areas are quiet and peaceful as Figure 7(f) shown. To conclude, Figure 7 gives the insights as follows:

- (1) From the state sequence heatmap, we notice that the state sequences are aligned by day, which indicates that the dynamics in the city have a period of one day, as the states in the same time slot but with different dates are usually the same.
- (2) The dynamic patterns within working days or no-working days are very similar, while the dynamics between working and non-working days are determined by the function of the region. For some regions like tourists and residence areas, they are similar like Figure 7(c) and (e), but for some educational and working places as Figure 7(b) and (f) express, the dynamics vary a lot.
- (3) Regions with similar functions are more likely to have similar dynamics, i.e., sharing more hidden states or having more similar transition patterns as Figure 7(a) and (b), which are both prestigious universities.

In summary, SSHMM not only detects the urban states with different activity level and modes but also reveal the rhythm of daily life.

5.3.3 Visualization of Urban Functions. To further evaluate our model to infer the functions, we cluster the state sequences by the method introduced in Section 4.4. We have empirically tried different value for the number of clusters as Figure 10 shown, which indicates that dividing dynamic patterns into 8 types is most suitable as the minimum DBI is achieved in this condition. We compare the result with the state-of-the-art functional zones discovering method, which is an LDA model by utilizing the static PoIs and mobility

together [28]. The geographical distribution of the regions with their function types is shown in Figure 11, where different colors present different functions. After clustering, we label their functions by the revealed semantic dynamics, including attractions, residence, business, education, company as well as some compound areas. We manually check some regions including the regions shown in Figure 7 on the map to verify their functions, which shows most of the regions with the same functions are divided into one cluster. Besides, clusters obtained by our model and LDA model have with the Normalized Mutual Information (NMI) of 0.25, which measures the similarity of two divisions with the range from -0.5 to 1 [14]. Therefore, the functions we infer from dynamic patterns are both validated by the manually labeled regions and the state-of-the-art method.

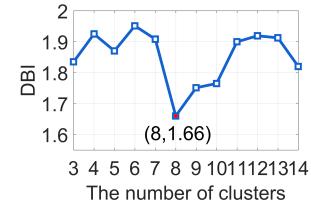


Figure 10: Cluster number identification.

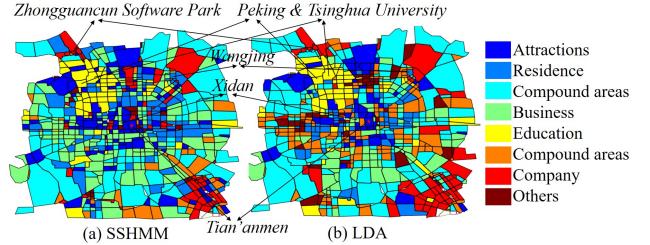


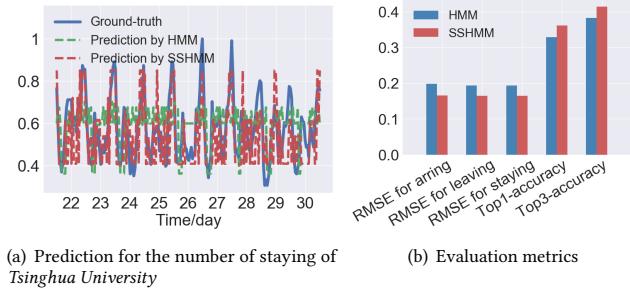
Figure 11: Visualization of the distribution of the regions with similar dynamics and functions.

As discussed above, the regions which have more common states and similar state transition processes are more likely to have the same functions. The result further demonstrates that our model is capable of inferring the distribution of functions all over the city.

5.3.4 Performance of Activity Prediction. In order to evaluate the prediction accuracy of our model, we first show the prediction of the number of staying for 9 days in *Tsinghua University* in Figure 12(a) as an example. Compared with the generate HMM, the recovered observations of our model are closer to the ground-truth in different time slots. The results of all the metrics are shown in Figure 12(b), where the average RMSE of population flow prediction is 0.195 and the Top3-accuracy for PoI popularity prediction is 41.4%, outperforming the HMM by 16% and 8%, respectively. These results demonstrate our SSHMM outperforms HMM in urban dynamics prediction problem.

6 IMPLICATIONS AND APPLICATIONS

Previously, most of the work w.r.t. understanding of urban dynamics came from traditional surveys conducted by human agents either physically or by phones [19]. While this way of collecting data provides detailed information about urban behaviors, it is hard to

**Figure 12: Prediction performance compared with HMM.**

update and present many weaknesses regarding generalization and scalability. In this paper, we propose a model to effectively and efficiently reveal urban dynamics via a large-scale mobile dataset. Regarding human activities as time series, we are able to detect typical urban states and different dynamic patterns. Compared with existing data-driven urban dynamics revealing systems [2, 18], we can model the aggregated activities in a concise and probabilistic way, which means dynamics prediction is also achieved at the same time. Besides, compared with previous work on urban function discovering which used human mobility data and statistic PoIs [25, 28], we are the first to utilize semantic check-ins and population-scale crowd flows as features to understand the dynamics and infer the relationship between dynamics and urban functions.

Our work has opened a new angle for urban representation. A modern city, as a comprehensive socio-economic environment, has a limited number of intrinsic states characterized by different activity levels and lifestyles, which can be learned from aggregated human mobility behaviors. As for urban temporal dynamics reflecting rhythms of daily life, they can be characterized by the transition on states. As a consequence, we are able to learn both the vector representation for both urban states (i.e., the mean of different feature dimension) and urban dynamics (i.e., the state sequence).

Our results facilitate urban planning in various aspects. Foremost, our findings provide a more comprehensive understanding of urban dynamics and urban functions. For architects, planners and urban designers, neighborhood activity patterns from intensive ethnographic surveys that take years to conduct can be outdated quickly given the rapid development of the society. Moreover, the features provided by our work helps to examine the intensity and diversity of human activity within a given neighborhood, thus offering novel insights into the functioning of the entire neighborhood and enabling the government to make better plan on land use. Last but not the least, understanding the regularity of the population flows are essential for traffic dispatching, transportation infrastructure construction, etc., while predicting the popular PoIs in different regions and different time slots will benefit for business siting and precision marking.

On the other hand, our proposed SSHMM, which learns a group HMMs tied by common state set, outperform a group of independent HMMs in the prediction task as it overcomes the data sparsity problem and uses the correlation between sequences. Similar to the aggregated mobility data, there is much other time series with the multi-dimensional nature and sheer size (e.g., voice sequence, music

sequence), which can utilize SSHMM for modeling and prediction. We leave to find more time series application as future work.

7 CONCLUSION

In this paper, we study the problem of understanding urban dynamics. We propose a State-sharing Hidden Markov Model (SSHMM), where all the regions share a common state set, but each region has its own transition regularity. To make it practical, we not only derive the inference of the parameters but also give an efficient and efficient algorithm to updating them. We evaluate our method via a real-life dataset in Beijing, which demonstrates that SSHMM learns semantics-rich urban dynamics model, recovers different activity regularities by a limited number of states and incurs low training cost. Our work opens a new angle to reveal and represent urban dynamics, and paves the way for extensive city applications.

APPENDIX

Based on Baum-Welch algorithm, the forward distribution $\alpha(s_{r,n})$ and backward distribution $\beta(s_{r,n})$ are first defined as follows,

$$\begin{aligned}\alpha(s_{r,n}) &= p(O_{r,n}|s_{r,n}) \sum_{s_{r,n-1}} \alpha(s_{r,n-1}) p(s_{r,n}|s_{r,n-1}), \\ \beta(s_{r,n}) &= \sum_{s_{r,n+1}} \beta(s_{r,n+1}) p(O_{r,n+1}|s_{r,n+1}) p(s_{r,n+1}|s_{r,n}),\end{aligned}\quad (15)$$

where $\alpha(s_{r,1}) = \pi_{r,k} p(O_{r,1}|s_{r,1} = k)$ and $\beta(s_{r,N} = k) = 1$. Then, two probability can be derived from α and β as follow,

$$\begin{aligned}\gamma(s_{r,n}) &= p(s_{r,n}|O_r) = \alpha(s_{r,n}) \beta(s_{r,n}) / p(O_r), \\ \xi(s_{r,n}, s_{r,n+1}) &= p(s_{r,n}, s_{r,n+1}|O_r) \\ &= \alpha(s_{r,n-1}) p(s_{r,n}|s_{r,n-1}) P(O_{r,n}|s_{r,n}) \beta(s_{r,n}) / p(O_r),\end{aligned}\quad (16)$$

where $p(O_r) = \sum_{s_{r,N}} \alpha(s_{r,n})$. Since the three terms in Equation 8 are independent, we can thus optimize the parameters π, A and $\{\mu, \sigma\}$ separately. To optimize π , we need to maximize the function namely,

$$f(\pi_{r,k}) = \sum_{r=1}^R \sum_S p(S|O_r, \theta_r^t) \ln \pi_{r,k} = \sum_{r=1}^R \sum_{k=1}^K \gamma(s_{r,1}^k) \ln \pi_{r,k}, \quad (17)$$

with the constraint $\sum_{k=1}^K \pi_{r,k} = 1$. Therefore, $\pi_{r,k}^{(t+1)} = \gamma(s_{r,1}^k)$. Similarly, to optimize A , we need to maximize the function namely,

$$\begin{aligned}h(A_{r,j,k}) &= \sum_{r=1}^R \sum_S \sum_{n=1}^{N-1} p(S|O_r, \theta_r^t) \ln p(s_{n+1}|s_n) \\ &= \sum_{r=1}^R \sum_{n=2}^N \sum_{j=1}^K \sum_{k=1}^K \xi(s_{r,n-1}^j, s_{r,n}^k) \ln A_{r,j,k},\end{aligned}\quad (18)$$

with the constraint $\sum_{k=1}^K A_{r,j,k} = 1$. Therefore, $A_{r,j,k}^{(t+1)} = \frac{1}{\Xi_j} \sum_{n=2}^N \xi(s_{r,n-1}^j, s_{r,n}^k)$.

To optimize $\{\mu, \sigma\}$, we need to maximize the function namely,

$$\begin{aligned}g(\mu, \sigma) &= \sum_{r=1}^R \sum_S \sum_{n=1}^N p(S|O_r, \theta_r^t) \ln p(O_{r,n}|\mu, \sigma) \\ &= \sum_{r=1}^R \sum_{n=1}^N \sum_{k=1}^K \gamma(s_{r,n}^k) \ln p(O_{r,n}|\mu, \sigma).\end{aligned}\quad (19)$$

Since $p(O_{r,n}|\mu, \sigma)$ is a continuous Gaussian function, when its derivative function equals 0, it gets the maximum value with μ, σ in Eq. 9.

REFERENCES

- [1] 2018. World urbanization prospects: The 2018 revision. United Nations Department of Economic and Social Affairs/Population Division. <https://drive.google.com/file/d/1FcNuCu6mlI6l5ZkmpWXZElelwPQKca4V/view?usp=sharing>.
- [2] Sofiane Abbar, Tahar Zanouda, Noora Al-Emadi, and Rachida Zegour. 2018. City of the People, for the People: Sensing Urban Dynamics via Social Media Interactions. In *Social Informatics*. 3–14.
- [3] J. R. Bellegarde and D. Nahamoo. 1990. Tied mixture continuous parameter modeling for speech recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing* 38, 12 (Dec 1990), 2033–2045. <https://doi.org/10.1109/29.61531>
- [4] Eunjoon Cho, Seth A. Myers, and Jure Leskovec. 2011. Friendship and Mobility: User Movement in Location-based Social Networks. In *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '11)*. ACM, 1082–1090.
- [5] D. L. Davies and D. W. Bouldin. 1979. A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1*, 2 (April 1979), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- [6] Arthur P Dempster, Nan M Laird, and Donald B Rubin. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)* (1977), 1–38.
- [7] Manuel GarcÃ¡a Docampo. 2014. Theories of Urban Dynamics. In *International Journal of Population Research*, Vol. 2014.
- [8] Fauziya F and Nijhawan G. 2014. A Comparative study of phoneme recognition using GMM-HMM and ANN based acoustic modeling. *International Journal of Computer Applications* 98, 6 (2014).
- [9] Marta C Gonzalez, Cesar A Hidalgo, and Barabasi. 2008. Understanding individual human mobility patterns. *Nature* 453, 779 (2008).
- [10] Xuedong Huang, Fileno Alleva, Hsiao-Wuen Hon, Mei-Yuh Hwang, Kai-Fu Lee, and Ronald Rosenfeld. 1993. The SPHINX-II speech recognition system: an overview. *Computer Speech & Language* 7, 2 (1993), 137 – 148. <http://www.sciencedirect.com/science/article/pii/S0885230883710077>
- [11] X. D. Huang. 1992. Phoneme classification using semicontinuous hidden Markov models. *IEEE Transactions on Signal Processing* 40, 5 (May 1992), 1062–1067. <https://doi.org/10.1109/78.134469>
- [12] Mei-Yuh Hwang and Xuedong Huang. 1993. Shared-Distribution Hidden Markov Models for Speech Recognition. 1 (11 1993), 414 – 420.
- [13] Xin Jin and Jiawei Han. 2011. K-medoids clustering. In *Encyclopedia of Machine Learning*. Springer, 564–565.
- [14] Zeger F Knops, JB Antoine Maintz, Max A Viergever, and Josien PW Pluim. 2006. Normalized mutual information based registration using k-means clustering and shading correction. *Medical image analysis* 10, 3 (2006), 432–439.
- [15] Kenneth Rose Liang Gu, Jayanth Nayak. 2000. Discriminative training of tied-mixture HMM by deterministic annealing. *ICSLP-2000* 4 (2000).
- [16] Ping Wang Marcus Berliant. 2015. Dynamic Urban Models: Agglomeration and Growth. In *Urban Dynamics and Growth: Advances in Urban Economics* (Mar 2015), 533–581. [https://doi.org/10.1108/S0573-8555\(2005\)0000266018](https://doi.org/10.1108/S0573-8555(2005)0000266018)
- [17] Wesley Mathew, Ruben Raposo, and Bruno Martins. 2012. Predicting future locations with hidden Markov models. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 911–918.
- [18] F. Miranda, H. Doraiswamy, M. Lage, K. Zhao, B. GonÃ‡alves, L. Wilson, M. Hsieh, and C. T. Silva. 2017. Urban Pulse: Capturing the Rhythm of Cities. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 791–800.
- [19] Jeffrey D Morenoff, Robert J Sampson, and Stephen W Raudenbush. 2001. Neighborhood inequality, collective efficacy, and the spatial dynamics of urban violence. *Criminology* 39, 3 (2001), 517–558.
- [20] Jiaul H. Paik. 2013. A Novel TF-IDF Weighting Scheme for Effective Ranking. In *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '13)*. ACM, 343–352.
- [21] L. R. Rabiner. 1989. A tutorial on hidden Markov models and selected applications in speech recognition. *Proc. IEEE* 77, 2 (Feb 1989), 257–286.
- [22] A. Viterbi. 1967. Error bounds for convolutional codes and an asymptotically optimum decoding algorithm. *IEEE Transactions on Information Theory* 13, 2 (April 1967), 260–269. <https://doi.org/10.1109/TIT.1967.1054010>
- [23] Lloyd R Welch. 2003. Hidden Markov models and the Baum-Welch algorithm. *IEEE Information Theory Society Newsletter* 53, 4 (2003), 10–13.
- [24] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Fuming Sun, and Fanchao Meng. 2018. Detecting Popular Temporal Modes in Population-scale Unlabelled Trajectory Data. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 2, 1, Article 46 (March 2018), 25 pages. <https://doi.org/10.1145/3191778>
- [25] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. 2018. Representing Urban Functions through Zone Embedding with Human Mobility Patterns. (2018), 3919–3925.
- [26] Donghan Yu, Yong Li, Fengli Xu, Pengyu Zhang, and Vassilis Kostakos. 2018. Smartphone App Usage Prediction Using Points of Interest. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 174 (Jan. 2018), 21 pages.
- [27] Jing Yuan, Yu Zheng, and Xing Xie. 2012. Discovering Regions of Different Functions in a City Using Human Mobility and POIs. In *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '12)*. ACM, 186–194. <https://doi.org/10.1145/2339530.2339561>
- [28] Nicholas Jing Yuan, Yu Zheng, Xing Xie, Yingzi Wang, Kai Zheng, and Hui Xiong. 2015. Discovering urban functional zones using latent activity trajectories. *IEEE Transactions on Knowledge and Data Engineering* 27, 3 (2015), 712–725.
- [29] Chao Zhang, Keyang Zhang, Quan Yuan, Haoruo Peng, Yu Zheng, Tim Hanratty, Shaowen Wang, and Jiawei Han. 2017. Regions, Periods, Activities: Uncovering Urban Dynamics via Cross-Modal Representation Learning. In *Proceedings of the 26th International Conference on World Wide Web (WWW '17)*. 361–370.
- [30] Chao Zhang, Keyang Zhang, Quan Yuan, Fangbo Tao, Luming Zhang, Tim Hanratty, and Jiawei Han. 2017. React: Online multimodal embedding for recency-aware spatiotemporal activity modeling. In *Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval*. ACM, 245–254.
- [31] Chao Zhang, Keyang Zhang, Quan Yuan, Luming Zhang, Tim Hanratty, and Jiawei Han. 2016. GMove: Group-Level Mobility Modeling Using Geo-Tagged Social Media. In *Proceedings of the 22Nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*. ACM, 1305–1314.
- [32] Junbo Zhang, Yu Zheng, and Dekang Qi. 2017. Deep Spatio-Temporal Residual Networks for Citywide Crowd Flows Prediction.. In *AAAI*. 1655–1661.
- [33] Ke Zhang, Qiuye Jin, Konstantinos Pelechrinis, and Theodoros Lappas. 2013. On the Importance of Temporal Dynamics in Modeling Urban Activity. In *Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing (UrbComp '13)*. ACM, 7:1–7:8. <https://doi.org/10.1145/2505821.2505825>
- [34] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. 2014. Urban Computing: Concepts, Methodologies, and Applications. *ACM Trans. Intell. Syst. Technol.* 5, 3, Article 38 (Sept. 2014), 55 pages. <https://doi.org/10.1145/2629592>
- [35] Xiao Zhou, Anastasios Noulas, Cecilia Mascolo, and Zhongxiang Zhao. 2018. Discovering Latent Patterns of Urban Cultural Interactions in WeChat for Modern City Planning. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery (KDD '18)*. ACM, 1069–1078. <https://doi.org/10.1145/3219819.3219929>
- [36] Wanzheng Zhu, Chao Zhang, Shuochao Yao, Xiaobin Gao, and Jiawei Han. 2018. A Spherical Hidden Markov Model for Semantics-Rich Human Mobility Modeling.. In *AAAI*.